

# LIMS and Data Pipelines

Fox Chase Cancer Center



# Why a LIMS

- Data Integrity
  - Protocol management (what was done)
  - Secure storage (what resulted)
  - Integrated QA/QC (how reliable)
- Data Communication
  - Electronic storage
  - APIs

# FCCC LIMS Projects

- Existing LIMS for Flow Cytometry
  - Web based Experimental Design
  - XML Communication to BD FACSVantage
  - Manual Web Link to BD FACScan (or other)
- Developing LIMS for Proteomics
  - Based on caLIMS
  - Initial Workflow and Data Schema Completed

# FCCC Data Pipelines

- **FGDP** (*Bioinformatics*, **20**, 282, 2004)
  - Web based microarray analysis
  - Automated simultaneous analyses
  - Java component architecture
- **ASAP** (*Bioinformatics*, **19**, 675, 2003)
  - Automated retrieval and piping of web data
  - mySQL with automated Perl script generation

# Development Plan

- Proteomics LIMS
  - Deploy initial database (done)
  - Work with researchers to provide minimally invasive integration with workflow
  - Refine through user feedback
- Links to Data and Annotation Pipelines
  - Automated QC/QA through pipeline
  - Automated updating of annotations

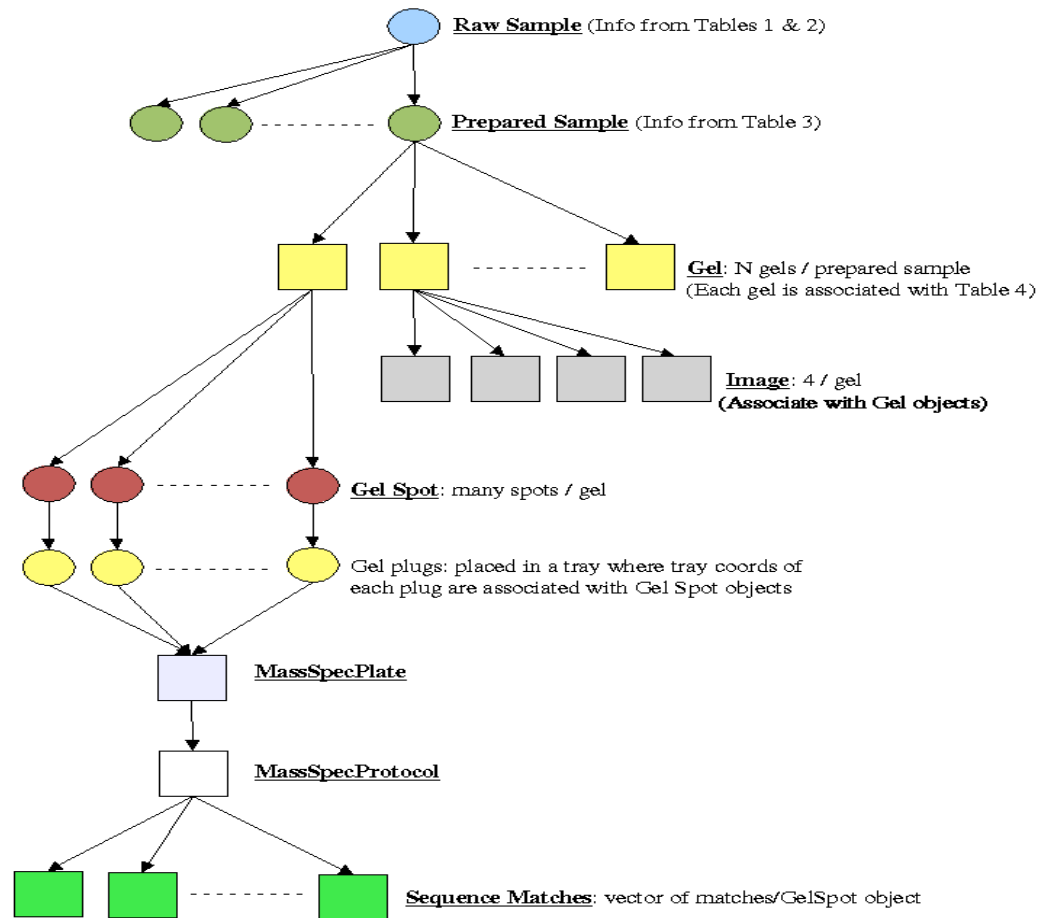
# Component Architecture

- LIMS
  - JSP for web interactions
  - JDBC for database connectivity
  - Multiplatform, multidatabase development
- FGDP
  - OO design patterns, easy module creation
  - Open source, share modules or package
  - Implements several TIGR MEV modules

# Architecture and Vocabularies

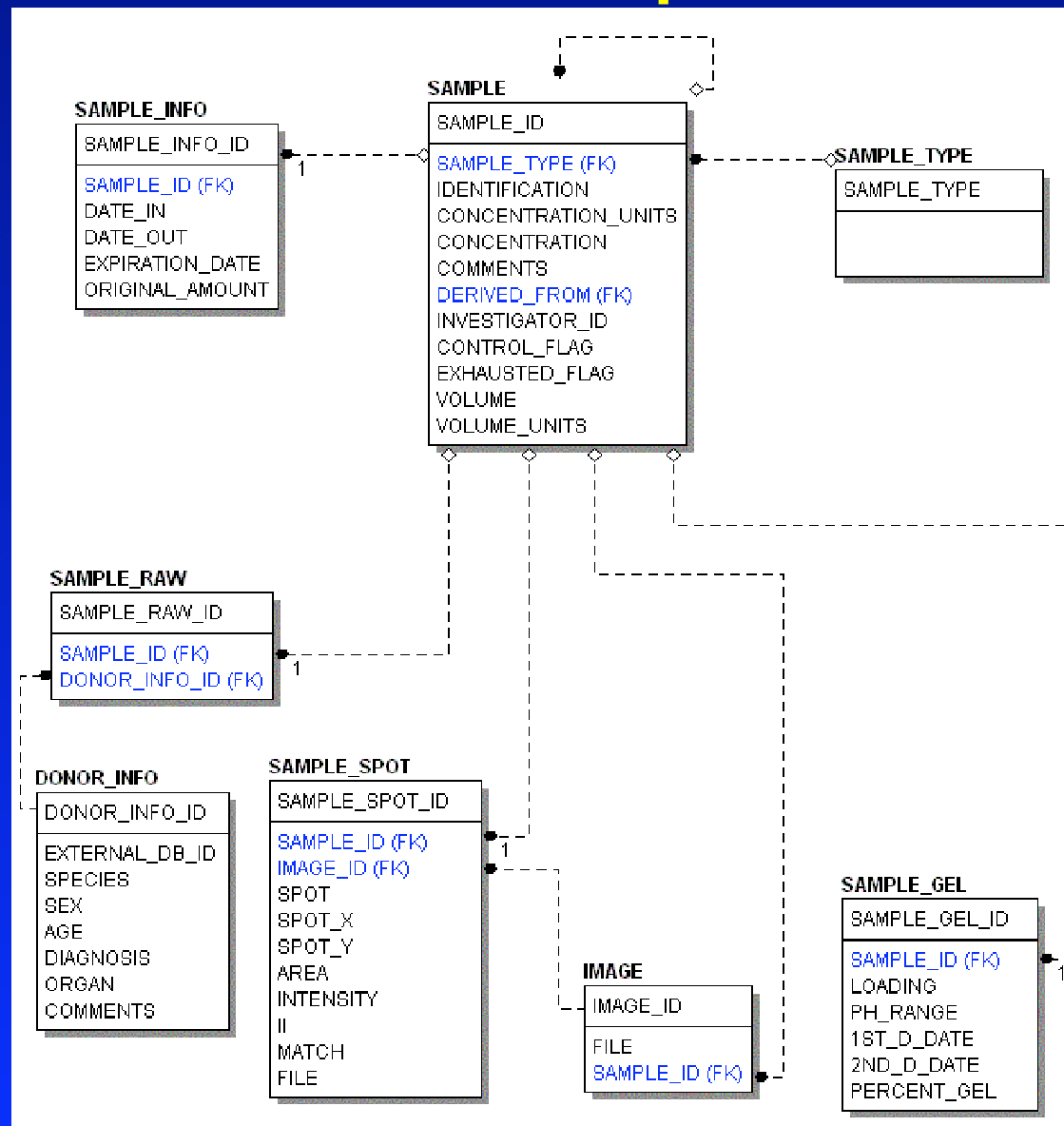
- Linking Systems and Data
  - XML and APIs for data sharing
  - Enable stable links between systems
- Identified Need for Vocabulary
  - Desire to link basic and clinical data (melanoma cell line to melanoma tumor)
  - Exploring limited vocabularies now
  - Need input from Vocabulary Group

# protLIMS Data Flow





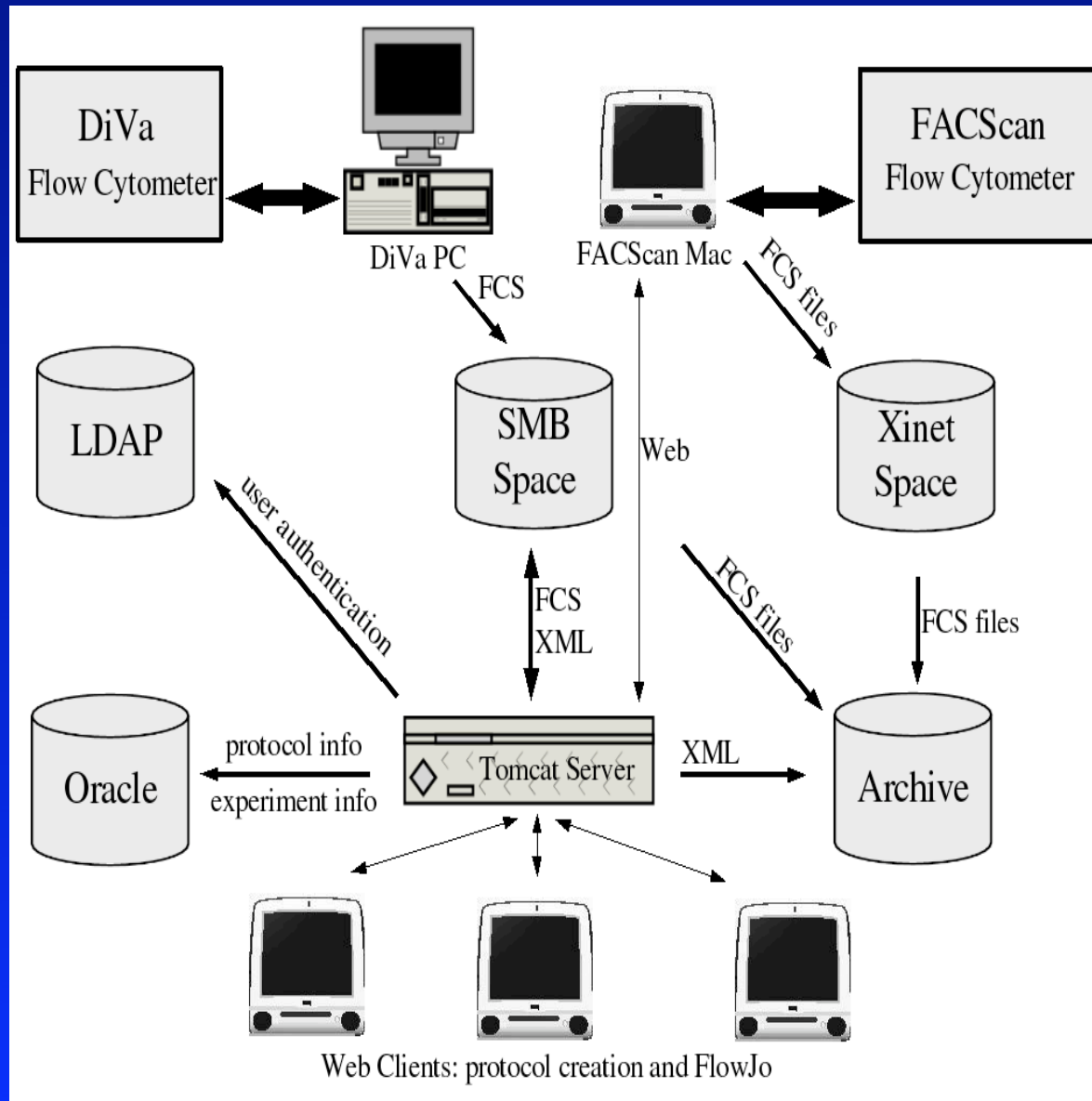
# protLIMS Sample Schema



# Development Plan

- Update Schema and Workflow
  - Work with adopters to generalize LIMS
  - Plan for additional data types
- Develop Interface
  - Work with adopters and FCCC researchers
  - Iterative development through matched development and production servers
- Example of flowLIMS

# flowLIMS



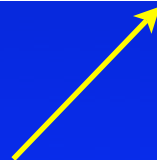
# Interface

Project Page  
after LDAP Login



The screenshot shows the 'Flow Cytometry LIMS' interface. At the top left is the 'FOX CHASE CANCER CENTER' logo. The title 'Flow Cytometry LIMS' is in large blue letters. On the top right are 'Project Level' and 'Help' buttons. Below the title, the user is identified as 'User: Ochs, Dr. Michael' with a 'Logout' button. The date 'Tue, Feb 17, 2004' is displayed on the right. A 'New Project:' section includes a text input field, a 'Create' button, and two search buttons: 'Search Protocol' and 'Search Experiment'. Below this, it says 'Showing results 1 through 2 out of 2 total' and a link 'Show Results in Blocks'. A table lists two projects: 'Workspace' and 'flowLIMS Demo', both owned by 'Ochs, Dr. Michael'. The 'Actions' column for 'Workspace' has a folder icon, while 'flowLIMS Demo' has a folder icon, a user icon, a 'no' icon, and a link icon.

Project Name	Project Owner	Colaborating Users	Actions
Workspace	Ochs, Dr. Michael		
flowLIMS Demo	Ochs, Dr. Michael		



Options include creating  
interactive, flexible groups

# Interface

Users set up  
cell types and stains  
that they routinely use

**FOX CHASE CANCER CENTER** **Flow Cytometry LIMS**

User: Ochs, Dr. Michael [Logout](#) [Back](#) [Project Level](#) [Help](#)

Tue, Feb 17, 2004

**Plate for test2**

**Cell editor**  
-Select-Cell-Type-  
BalbC  
Rag1KOBM  
Rag1KOd16FL

**Stain editor**  
-Select-Fl-Stain-  
FL \* B220  
FL \* CD3  
FL \* CD4  
FL \* Ly6c  
APC \* CD8  
Indo1B \* CD3

	1	2	3	4	5	6	7	8	9	10	11	12
A	A-1	A-2	A-3	A-4	A-5	A-6	A-7	A-8	A-9	A-10	A-11	A-12
B	B-1	B-2	B-3	B-4	B-5	B-6	B-7	B-8	B-9	B-10	B-11	B-12
C	C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10	C-11	C-12
D	D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9	D-10	D-11	D-12
E	E-1	E-2	E-3	E-4	E-5	E-6	E-7	E-8	E-9	E-10	E-11	E-12
F	F-1	F-2	F-3	F-4	F-5	F-6	F-7	F-8	F-9	F-10	F-11	F-12
G	G-1	G-2	G-3	G-4	G-5	G-6	G-7	G-8	G-9	G-10	G-11	G-12
H	H-1	H-2	H-3	H-4	H-5	H-6	H-7	H-8	H-9	H-10	H-11	H-12

**Stains:** -Sample-Fl-Stain- **Cell type:**

[Save](#) [Details](#) [Edit](#)

Virtual pipetting allows users  
to easily layout experiment

After pipetting, users  
can see what is in each well

# Interface

Experiments can be published or private

FOX CHASE  
CANCER CENTER

## Flow Cytometry LIMS

User: Ochs, Dr. Michael Logout

Tue, Feb 17, 2004

Search Protocol Search Experiment Copy to project... OK

Project:  
Showing results 1 through 10 out of 96 total [Show All Results](#)

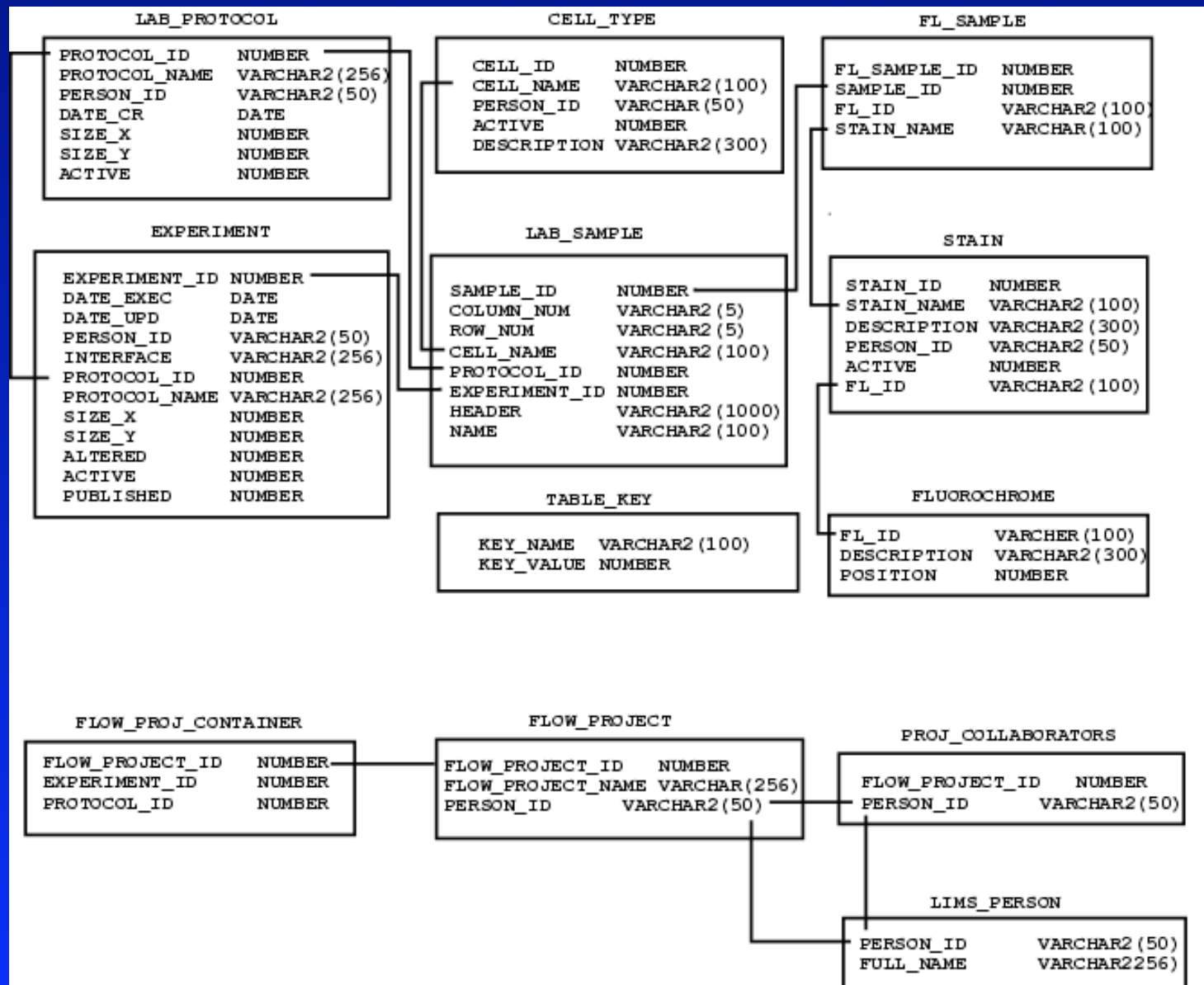
Check <a href="#">All</a>	Title	Owner	Date	Type	Exp Id	Pub	Action
<input type="checkbox"/>	Test Alt Compensation	Hardy, Randy	02/11/2004	experiment	1724	✓	
<input type="checkbox"/>	N2ic-transfectants-for WB	Wen, Li-Jun	02/09/2004	experiment	1717	✓	
<input type="checkbox"/>	J558-012604PMIGII	Tang, Ling	02/06/2004	experiment	1712	✓	
<input type="checkbox"/>	B875(II)d5 BM Notch2GFP	Hayakawa, Kyoko	02/06/2004	experiment	1710	✓	
<input type="checkbox"/>	J558-012604PMIGII	Tang, Ling	02/02/2004	experiment	1683	✓	
<input type="checkbox"/>	MuKTest_012904	Rowley, Benjamin	01/29/2004	experiment	1662	✓	
<input type="checkbox"/>	PhosStain_010204	Rowley, Benjamin	01/27/2004	experiment	1643	✓	
<input type="checkbox"/>	MuTest_012704	Rowley, Benjamin	01/27/2004	experiment	1645	✓	
<input type="checkbox"/>	KTest_012704	Rowley, Benjamin	01/27/2004	experiment	1646	✓	
<input type="checkbox"/>	J558-012604PMIGII	Tang, Ling	01/26/2004	experiment	1638	✓	

[First](#) [Previous](#) [Next](#) [Last](#)

Automatic generation and  
downloading of  
FlowJo Summary Files

Additional files can  
be added to system

# flowLIMS Schema



# FGDP

- Automated, multiple, simultaneous analyses of functional genomics data
- Java based using RMI for distribution
- Plan to use for QA/QC as well as data analysis for Proteomics LIMS



# Models Data Flow

Parsing/Image Analysis

Normalization

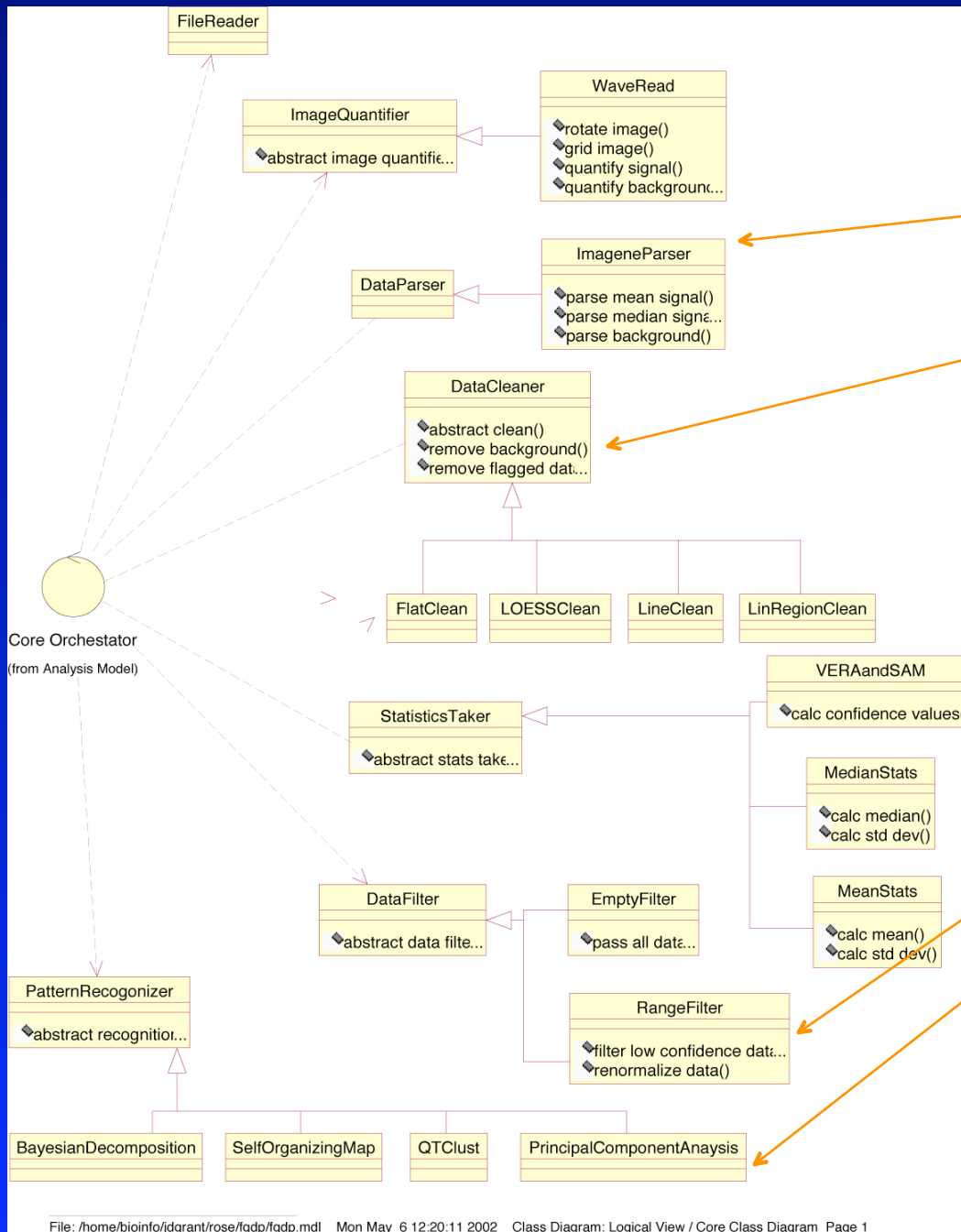
Statistical Analysis

Filtering

Pattern Recognition

Image Generation

Output Parsing



# functional genomics data pipeline

[Results](#)
[Help](#)
[Logout](#)

[1. Files](#)
[2. Modules](#)
[3. Filters](#)
[4. Options](#)
[5. Launch](#)

## Modules selected:

Data Cleanup	LOESS Fit [Module 1]	Params:	Keep Flagged <input checked="" type="radio"/> True <input type="radio"/> False Dim Point Mode <input type="text" value="Floor all"/> Floor factor <input type="text" value="1.6"/> <i>Enter a real number</i> Smoothing Parameter <input type="text" value="0.9"/> <i>Enter a real number</i>
Statistics	Basic Stats [Module 1]	Params:	Mean or Median <input type="text" value="Mean"/>
Statistics	VERA and SAM [Module 1]	Params:	Stopping Change Value <input type="text" value="0.001"/> <i>Enter a real number</i> Max Iterations <input type="text" value="500"/> <i>Enter an integer</i> Min Iterations <input type="text" value="300"/> <i>Enter an integer</i>
Clustering	Cast Clustf [Module 1]	Params:	DF <input type="text" value="Pearson Correlation Coefficient"/> threshold <input type="text" value="0.9"/>
Clustering	QTclustFCCC [Module 1]	Params:	DF <input type="text" value="Covariance"/> diameter <input type="text" value="10"/> Min Cluster Size <input type="text" value="10"/>
Clustering	K-Means [Module 1]	Params:	DF <input type="text" value="Pearson Correlation Coefficient"/> iterations <input type="text" value="200"/> clusters <input type="text" value="6"/> <i>Enter an integer</i>

Warning: If either **Filtering** or **Pattern Recognition** has no modules selected, neither of them will run.

[Verify](#)
[Reset](#)

◆◆◆ Pipelines: ◆◆◆

	Data Cleanup	Statistics	Filtering	Clustering																		
1	<table><tr><td>Module:</td><td>LOESS Fit</td></tr><tr><td>Params:</td><td>Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9</td></tr></table>	Module:	LOESS Fit	Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9	<table><tr><td>Module:</td><td>VERA and SAM</td></tr><tr><td>Params:</td><td>Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300</td></tr></table>	Module:	VERA and SAM	Params:	Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300	<table><tr><td colspan="2">myFilter</td></tr><tr><td>Module:</td><td>Params:</td></tr><tr><td>MinChange Filter</td><td>Factor: 1.1</td></tr></table>	myFilter		Module:	Params:	MinChange Filter	Factor: 1.1	<table><tr><td>Module:</td><td>Cast Clustf</td></tr><tr><td>Params:</td><td>DF: 1 threshold: 0.9</td></tr></table>	Module:	Cast Clustf	Params:	DF: 1 threshold: 0.9
	Module:	LOESS Fit																				
Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9																					
Module:	VERA and SAM																					
Params:	Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300																					
myFilter																						
Module:	Params:																					
MinChange Filter	Factor: 1.1																					
Module:	Cast Clustf																					
Params:	DF: 1 threshold: 0.9																					
2	<table><tr><td>Module:</td><td>LOESS Fit</td></tr><tr><td>Params:</td><td>Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9</td></tr></table>	Module:	LOESS Fit	Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9	<table><tr><td>Module:</td><td>VERA and SAM</td></tr><tr><td>Params:</td><td>Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300</td></tr></table>	Module:	VERA and SAM	Params:	Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300	<table><tr><td colspan="2">myFilter</td></tr><tr><td>Module:</td><td>Params:</td></tr><tr><td>MinChange Filter</td><td>Factor: 1.1</td></tr></table>	myFilter		Module:	Params:	MinChange Filter	Factor: 1.1	<table><tr><td>Module:</td><td>QTClustFCCC</td></tr><tr><td>Params:</td><td>DF: 3 diameter: 0.2 Min Cluster Size: 5</td></tr></table>	Module:	QTClustFCCC	Params:	DF: 3 diameter: 0.2 Min Cluster Size: 5
	Module:	LOESS Fit																				
Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9																					
Module:	VERA and SAM																					
Params:	Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300																					
myFilter																						
Module:	Params:																					
MinChange Filter	Factor: 1.1																					
Module:	QTClustFCCC																					
Params:	DF: 3 diameter: 0.2 Min Cluster Size: 5																					
3	<table><tr><td>Module:</td><td>LOESS Fit</td></tr><tr><td>Params:</td><td>Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9</td></tr></table>	Module:	LOESS Fit	Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9	<table><tr><td>Module:</td><td>VERA and SAM</td></tr><tr><td>Params:</td><td>Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300</td></tr></table>	Module:	VERA and SAM	Params:	Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300	<table><tr><td colspan="2">myFilter</td></tr><tr><td>Module:</td><td>Params:</td></tr><tr><td>MinChange Filter</td><td>Factor: 1.1</td></tr></table>	myFilter		Module:	Params:	MinChange Filter	Factor: 1.1	<table><tr><td>Module:</td><td>K-Means</td></tr><tr><td>Params:</td><td>DF: 1 iterations: 200 clusters: 6</td></tr></table>	Module:	K-Means	Params:	DF: 1 iterations: 200 clusters: 6
	Module:	LOESS Fit																				
Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9																					
Module:	VERA and SAM																					
Params:	Stopping Change Value: 0.001 Max Iterations: 500 Min Iterations: 300																					
myFilter																						
Module:	Params:																					
MinChange Filter	Factor: 1.1																					
Module:	K-Means																					
Params:	DF: 1 iterations: 200 clusters: 6																					
4	<table><tr><td>Module:</td><td>LOESS Fit</td></tr><tr><td>Params:</td><td>Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9</td></tr></table>	Module:	LOESS Fit	Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9	<table><tr><td>Module:</td><td>Basic Stats</td></tr><tr><td>Params:</td><td>Mean or Median: mean</td></tr></table>	Module:	Basic Stats	Params:	Mean or Median: mean	<table><tr><td colspan="2">myFilter</td></tr><tr><td>Module:</td><td>Params:</td></tr><tr><td>MinChange Filter</td><td>Factor: 1.1</td></tr></table>	myFilter		Module:	Params:	MinChange Filter	Factor: 1.1	<table><tr><td>Module:</td><td>Cast Clustf</td></tr><tr><td>Params:</td><td>DF: 1 threshold: 0.9</td></tr></table>	Module:	Cast Clustf	Params:	DF: 1 threshold: 0.9
	Module:	LOESS Fit																				
Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9																					
Module:	Basic Stats																					
Params:	Mean or Median: mean																					
myFilter																						
Module:	Params:																					
MinChange Filter	Factor: 1.1																					
Module:	Cast Clustf																					
Params:	DF: 1 threshold: 0.9																					
5	<table><tr><td>Module:</td><td>LOESS Fit</td></tr><tr><td>Params:</td><td>Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9</td></tr></table>	Module:	LOESS Fit	Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9	<table><tr><td>Module:</td><td>Basic Stats</td></tr><tr><td>Params:</td><td>Mean or Median: mean</td></tr></table>	Module:	Basic Stats	Params:	Mean or Median: mean	<table><tr><td colspan="2">myFilter</td></tr><tr><td>Module:</td><td>Params:</td></tr><tr><td>MinChange Filter</td><td>Factor: 1.1</td></tr></table>	myFilter		Module:	Params:	MinChange Filter	Factor: 1.1	<table><tr><td>Module:</td><td>QTClustFCCC</td></tr><tr><td>Params:</td><td>DF: 3 diameter: 0.2 Min Cluster Size: 5</td></tr></table>	Module:	QTClustFCCC	Params:	DF: 3 diameter: 0.2 Min Cluster Size: 5
	Module:	LOESS Fit																				
Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9																					
Module:	Basic Stats																					
Params:	Mean or Median: mean																					
myFilter																						
Module:	Params:																					
MinChange Filter	Factor: 1.1																					
Module:	QTClustFCCC																					
Params:	DF: 3 diameter: 0.2 Min Cluster Size: 5																					
6	<table><tr><td>Module:</td><td>LOESS Fit</td></tr><tr><td>Params:</td><td>Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9</td></tr></table>	Module:	LOESS Fit	Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9	<table><tr><td>Module:</td><td>Basic Stats</td></tr><tr><td>Params:</td><td>Mean or Median: mean</td></tr></table>	Module:	Basic Stats	Params:	Mean or Median: mean	<table><tr><td colspan="2">myFilter</td></tr><tr><td>Module:</td><td>Params:</td></tr><tr><td>MinChange Filter</td><td>Factor: 1.1</td></tr></table>	myFilter		Module:	Params:	MinChange Filter	Factor: 1.1	<table><tr><td>Module:</td><td>K-Means</td></tr><tr><td>Params:</td><td>DF: 1 iterations: 200 clusters: 6</td></tr></table>	Module:	K-Means	Params:	DF: 1 iterations: 200 clusters: 6
	Module:	LOESS Fit																				
Params:	Keep Flagged: true Dim Point Mode: 1 Floor factor: 1.6 Smoothing Parameter: 0.9																					
Module:	Basic Stats																					
Params:	Mean or Median: mean																					
myFilter																						
Module:	Params:																					
MinChange Filter	Factor: 1.1																					
Module:	K-Means																					
Params:	DF: 1 iterations: 200 clusters: 6																					
	Data Cleanup	Statistics	Filtering	Clustering																		

Run

# LIMS and Pipelines

- Open Source (Components)
  - Either GNU GPL or LPL
  - Full Documentation and Support
- Integration with Adopters
  - Feedback required for Development
  - Group Expertise in Science and Software
- Integration with Working Groups
  - Vocabularies
  - Architecture

# Development Team

- Jeffrey Grant, Architect
- Yue Zhang, Sr. Programmer
- Elizabeth Goralczyk, Sr. Prog/Analyst
- Michael Slifker, Sr. Prog/Analyst
- Luke Somers, Prog/Analyst
- Olga Tchuvatkina, Prog/Analyst